

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
імені ТАРАСА ШЕВЧЕНКА
Факультет інформаційних технологій
Кафедра прикладних інформаційних систем

122 «Комп'ютерні науки»
(шифр і назва спеціальності)

«Прикладне програмування»
(назва освітньої програми)

Кваліфікаційна робота бакалавра

на тему: «Прикладна програма пошуку резюме кандидатів за користувацькими критеріями»



Виконав

(Підпис)

Войналович Андрій Олександрович
(прізвище, ім'я, по батькові)

Керівник Сайко Володимир Григорович
(прізвище, ім'я, по батькові)



_____ **Попередній захист:**

_____ *до захисту*

(Висновок: “До захисту в екзаменаційній комісії”)

Завідувач кафедри _____ Плескач В.Л.
(Дата) (Підпис) (Прізвище, ініціали)

Засвідчую, що у цій дипломній роботі немає запозичень із праць інших авторів без відповідних посилань. Унікальність тексту - 81 %

Київ – 2022

Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій

Кафедра прикладних інформаційних систем

Назва теми: «Прикладна програма пошуку резюме кандидатів за користувацькими критеріями»

Освітня програма: Прикладне програмування

Спеціальність: Комп'ютерні науки

ПІБ

Підпис

Войналович Андрій Олександрович



Назва роботи українською та англійською мовами

Прикладна програма пошуку резюме кандидатів за користувацькими критеріями

Software system for automation of personnel selection.

Мета кваліфікаційної роботи бакалавра, завдання

Мета роботи: Пошук та агрегація переліку резюме користувачів популярних українських ресурсів для пошуку роботи, ґрунтуючись на заданих у програму даних.

План роботи:

1. Підходи до розроблення сучасних застосунків пошуку кандидатів
2. Аналіз програмно-технологічних рішень пошуку кандидатів
3. Реалізація та впровадження застосунку пошуку та агрегації кандидатів за заданими параметрами

ПІБ, ступінь, звання наукового керівника роботи:



КАЛЕНДАРНИЙ ПЛАН ВИКОНАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ БАКАЛАВРА

Номер	Назва етапів кваліфікаційної роботи бакалавра	Термін виконання етапів кваліфікаційної роботи	Відмітка про виконання
	Вибір теми та наукового керівника дипломної роботи	23.10.2021	
	Видача завдання дипломної роботи	02.11.2021	
	Настановча групова співбесіда з дипломної роботи	01.12.2021	
	Затвердження плану дипломної роботи	21.12.2021	
	Підбір та вивчення літературних та інших джерел з теми дослідження	29.01.2022	
	Підготовка і подання науковому керівнику першого варіанту I розділу роботи	02.02.2022	
	Підготовка і подання науковому керівнику першого варіанту II розділу роботи	22.02.2022	
	Підготовка і подання науковому керівнику першого варіанту III розділу роботи	15.03.2022	
	Подання роботи у першому варіанті	29.03.2022	
	Оформлення пояснювальної записки дипломної роботи	26.04.2022	
	Подання дипломної роботи на попередній захист	20.05.2022	
	Врачування зауважень керівника і подання роботи в остаточному варіанті (з відповідним висновком про допуск) на кафедрі	01.06.2022	
13.	Затвердження роботи в цілому (підготовка письмового відгуку керівника, письмова рецензія на кваліфікаційну роботу бакалавра)	15.06.2022	
14.	Захист дипломної роботи	24.06.2022	

Здобувач вищої освіти



(підпис)

Керівник _____
(підпис)

ВІДОМІСТЬ ДИПЛОМНОЇ РОБОТИ

Складові частини дипломної роботи	Обсяг, арк.
Титульний аркуш	1
Календарний план дипломної роботи	1
Відомість дипломної роботи	1
Анотація	1
Анотація (іноземною мовою-англійською)	1
Зміст	1
Перелік скорочень, умовних позначень, термінів	1
Вступ	3
1	12
2	6
3	15
Висновки	1
Перелік використаних джерел	2
Додатки	0

				ДП ХХХХ 00.000.00			
	ПІБ	Підп.	Дата	Відомість дипломної роботи	Лис т	Листів	
Розробн.	Войналович А.О.						
Керівн.	Сайко В.Г.						
Н/контр.	Базилук А.М.						
Зав.каф.	Плескач В.Л.						

АНОТАЦІЯ

Дипломна робота: 45 с., 12 рис., 3 табл., 11 джерел.

Ця дипломна робота присвячена проектуванню та розробці прикладного програмного забезпечення для пошуку резюме за критеріями користувача.

Метою дипломної роботи є розроблення програмного забезпечення для отримання та структурування даних резюме з відкритих джерел, за користувацькими критеріями, для підвищення процесу взаємодії роботодавців з майбутнім персоналом компаній.

Завдання дослідження:

Для досягнення поставленої мети треба вирішити такі **завдання**:

- Отримати теоретичні засади щодо розроблення програмного забезпечення для отримання даних резюме з відкритих джерел
- Розробити програмний продукт пошуку резюме за критеріями користувача вимог та з урахуванням попередніх досліджень.

Об'єкт дослідження.

Процеси створення програмного забезпечення для отримання та структурування даних резюме з відкритих джерел, за користувацькими критеріями.

Предмет дослідження.

Програмні засади підвищення ефективності процесу взаємодії роботодавців з майбутнім персоналом компаній.

Методи дослідження.

Аналіз сучасних існуючих програмних рішень пошуку резюме кандидатів за користувацькими критеріями.

Аналіз стандартів та практик створення програмних рішень з пошуку резюме кандидатів за вказаними користувацькими критеріями використовуючи сучасну технологію Web Scraping.

Ключові слова: автоматизація, пошук кадрів, резюме, веб-скрейпінг.

ABSTRACT

Thesis: 45 p., 12 fig., 3 table., 11 sources.

This thesis is devoted to the design and development of application software to search for resumes according to user criteria.

The purpose of the thesis is to develop software for obtaining and structuring resume data from open sources, according to user criteria, to increase the process of interaction of employers with future company personnel.

Research objectives:

To achieve this goal, it is necessary to solve the following **tasks**:

1. Get theoretical foundations for the development of software for obtaining resume data from open sources
2. Develop a software product search resume according to user requirements and considering previous research.

Object of research.

The processes of creating software for obtaining and structuring resume data from open sources, according to user criteria.

Subject of research.

Program principles for improving the efficiency of the process of interaction between employers and future company personnel.

Research methods.

Analysis of modern existing software solutions for searching for resumes of candidates according to user criteria.

Analysis of standards and practices for creating software solutions for finding resumes of candidates according to these user criteria using modern Web Scraping technology.

Keywords: automation, frame search, resume, web scraping.

Зміст

ВСТУП	9
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ПОБУДОВИ ПРОГРАМ ЗВАЄМОДІЇ ІЗ КОРИСТУВАЦЬКИМ ІНТЕРФЕЙСОМ	12
1.1 Поняття програмних систем фільтрації, агрегації та репрезентації резюме що ґрунтуються на даних із відкритих джерел	12
1.2 Поняття та історія створення технології автоматизації бізнес-процесів компанії із використанням технології RPA	13
1.3 Технологія Web Scraping та її різновиди для впровадження системи пошуку резюме за даними із відкритих архівів	15
1.4 Історія створення програмних рішень систем фільтрації, агрегації та репрезентації резюме	18
1.5 Міжнародний досвід на прикладі систем агрегації резюме	20
РОЗДІЛ 2. АНАЛІЗ ПІДХОДІВ І ПРИНЦИПІВ ПОБУДОВИ ПРИКЛАДНОГО ЗАСТОСУНКУ ПОШУКУ РЕЗЮМЕ	24
2.1 Аналіз принципів побудов застосунків автоматизації бізнес-процесів, використовуючи технологію RPA	24
2.2 Технології RPA та DataScraping як основне архітектурне рішення для побудови застосунку	25
2.3 Різні види забезпечення побудови прикладного застосунку із пошуку резюме за джерелом відкритих даних	27
РОЗДІЛ 3. ПРОЕКТУВАННЯ, РОЗРОБЛЕННЯ ТА ВПРОВАДЖЕННЯ ПРОГРАМНОЇ СИСТЕМИ ПОШУКУ РЕЗЮМЕ	30
3.1 Опис концепції та програмної реалізації технології Data Scraping при розробці прикладної системи	30
3.2 Технічне завдання та постановка задачі програми пошуку та структурування даних резюме із відкритих джерел	34
3.3 Опис і склад програмних модулів програмної системи автоматизації пошуку резюме за визначеними користувацькими критеріями.	36
3.4 Стек використовуваних ІТ, впровадження системи пошуку резюме за відкритими джерелами	39
3.5 Інструкції користувача, адміністратора програми пошуку та структурування даних резюме із відкритих джерел	41
ВИСНОВКИ	45
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	46

ВСТУП

Наразі питання оновлення кадрового складу стали актуальними для багатьох організацій, у результаті чого постійний пошук нових кандидатів, формування кадрового резерву, удосконалення підбору працівників стали найголовнішим завданням фахівців із управління персоналом. Змінний ринок праці вимагає постійного розвитку професійних навичок та компетенцій персоналу, що впливає на зміни у вимогах до нових співробітників організації та ускладнення технологій підбору персоналу. Через високу частоту зміни кадрів багато організацій потребують постійного поповнення персоналу при мінімальних витратах на процесі підбору, у який входять наступні кроки: опис та розміщення вакансії,

Залучення релевантних кандидатів, їх оцінка та відбір, наймання в організацію кандидатів із найбільш високим рівнем підготовки та релевантності до конкретних задач. Це завдання допомагають вирішувати технології автоматизації підбору персоналу. За останні кілька років сфера управління персоналом пройшла великий шлях у сфері автоматизації кадрових процесів. Одним із перших процесів, який було успішно придано автоматизації – це пошук та відбір кандидатів на вакантне місце за певними критеріями. Традиційні методи підбирання із великою частотою висвітлюються у науковій літературі, та методи з використанням цифрових технологій активно пішли в область практичного застосування у провідних ІТ-компаніях світу. Можна виділити низку технологій автоматизації процесів підбору персоналу, що мають найбільший інтерес у практичній роботі фахівців сфері управління персоналом:

- використання чат-ботів для контакту із потенційним працівником;
- інформування про вакансії за допомогою QR-кодів;
- реалізація відеоінтерв'ю;
- розробка кар'єрного сайту;
- пошук персоналу на спеціальних сайтах-агрегаторах;
- пошук персоналу через соціальні мережі (SMM);

Автоматизація підбору працівників дозволяє:

- скоротити трудовитрати, пов'язані з плануванням та реалізацією підбору кадрів;
- Мінімізувати витрати на підбір кадрів;
- прискорити пошук та залучення кандидатів;
- Забезпечити постійне зростання бази потенційних кандидатів;
- здійснити грамотний облік претендентів та виключити втрату даних;
- проводити якісне, повноцінне опрацювання кожного резюме;
- збільшити зростання конверсії потенційних кандидатів, претендентів на працевлаштованих співробітників, працівників;
- забезпечити оптимізацію управління HR-напрямком компанії

Майже двадцять років збору загальнодоступних даних є ключовою основою для багатьох підприємств у багатьох секторах. Це тому, що Інтернет став найбільшим ресурсом даних у світі, а знання для бізнесу більше не надходять лише із застарілих каналів, таких як звіти та ручні бази даних, а також із майже реальних даних з Інтернету. Технологія веб-скрейпінг, що стала доступна будь-якому користувачу ПК дозволяє керівникам приймати більш обґрунтовані рішення, які сильно впливають на їхні організаційні та операційні стратегії, а також на результати бізнесу.

Існує багато переконливих прикладів використання в академічних та бізнесових сферах, які підкреслюють важливість збору та аналізу загальнодоступних веб-даних. Наприклад, провідні компанії використовують цю технологію для збору інформації про стан ринків, аналітичних даних про конкурентів, таких як ціни та рівень запасів, а також настрої споживачів. Дослідники, науковці, інвестори та журналісти також використовують публічний веб-скрейпінг у своїх стратегіях даних, щоб отримати уявлення в реальному часі та заснувати свої звіти на надійних даних. Вони включають погляд на суспільні настрої та добробут,

структуру організаційної команди, перспективи зростання та конкурентний ландшафт для залучення цільової аудиторії.

Отже, система автоматизації дозволяє скоротити часові витрати, пов'язані з процесом пошуку, сортування та найму персоналу. Багаторазово спрощуються використовувані операції компанії (збір та обробка резюме, листування з кандидатами та клієнтами, формування єдиної бази, поточне узгодження, аналіз подій тощо), знижується навантаження на HR-фахівців. У даній роботі буде розглянуто можливість створення прикладного застосунку для агрегації резюме із декількох провідних українських інтернет ресурсів.

Метою дипломної роботи є розроблення програмного забезпечення для отримання та структурування даних резюме з відкритих джерел, за користувацькими критеріями, для підвищення процесу взаємодії роботодавців з майбутнім персоналом компаній.

Основним призначенням практики є вивчення теоретичних основ створення програмних рішень для пошуку резюме за певними користувацькими критеріями, а також набуття практичного досвіду по роботі з продуктом на якому ґрунтується побудова програм для взаємодії із користувацьким інтерфейсом архівів - UiPath.

Цілі практики були досягнуті завдяки вирішенню наступних задач:

- засвоєння основних положень про методи пошуку найрелевантніших вакансій ринку використовуючи ресурси відкритих даних;
- вивчення функціональних можливостей середовища розробки UiPath;
- розробка програми яка виконує пошук найрелевантніших резюме за користувацькими параметрами, що було задано у зручному для користувача форматі.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ПОБУДОВИ ПРОГРАМ ВЗАЄМОДІЇ ІЗ КОРИСТУВАЦЬКИМ ІНТЕРФЕЙСОМ

1.1 Поняття програмних систем фільтрації, агрегації та репрезентації резюме що ґрунтуються на даних із відкритих джерел

У галузі програмних систем агрегатор резюме, який також називають агрегатором каналів, читачем каналів, читачем новин, читачем RSS або просто агрегатором, є клієнтським програмним забезпеченням або веб-додатком, який об'єднує структурований веб-контент, такий як інтернет-газети, блоги, подкасти та відео-блоги (vlogs) в одному місці для зручного перегляду.

Часте відвідування багатьох окремих веб-сайтів, щоб з'ясувати, чи оновлено вміст сайту, займає багато часу тому технологія агрегації допомагає об'єднати декілька веб-сайтів в одну сторінку, яка може відображати лише нову або оновлену інформацію з багатьох сайтів. Агрегатори скорочують час та зусилля необхідні для регулярної перевірки веб-сайтів на наявність оновлень, створюючи унікальний інформаційний простір або особистий структурований перелік новин. Після підписки на канал, агрегатор може перевіряти наявність нового вмісту через визначені користувачем інтервали та отримувати оновлення. На відміну від одержувачів деякої push-інформації, користувач агрегатора може скасувати підписку на стрічку.

Архівні джерела із даних резюме шукачів часто містяться у форматах RSS або Atom, які використовують розширювану мову розмітки (XML) для структурування частин інформації, що підлягають агрегуванню у програмі зчитування контенту, яка відображає інформацію у зручному для користувача інтерфейсі. Перш ніж підписатись розсилку нових резюме-файлів шукачів, користувачі повинні встановити або програми «перегляд останніх доданих резюме», або «агрегатор резюме», щоб прочитати її. Агрегатор забезпечує консолідований перегляд вмісту в одному дисплеї браузера або настільній програмі.

"Настільні програми пропонують переваги потенційно розширеного користувацького інтерфейсу та можливості надання певного вмісту, навіть коли комп'ютер не підключений до Інтернету. Інтернет-читачі каналів пропонують велику зручність, дозволяючи користувачам отримувати доступ до сучасних стрічок з будь-який комп'ютер, підключений до Інтернету".

Хоча деякі програми матимуть автоматизований процес передплати на стрічку новин, основним способом підписки є натискання на піктограму веб-стрічки та / або текстове посилання.

Функції агрегації часто вбудовуються на веб-сайти веб-порталів, у самі веб-браузери, в програми електронної пошти або в прикладне програмне забезпечення, розроблене спеціально для читання каналів. Агрегатори з можливостями подкастингу можуть автоматично завантажувати мультимедійні файли, такі як записи MP3.

У деяких випадках їх можна автоматично завантажити на портативні медіаплеєри (наприклад, iPod), коли вони підключені до комп'ютера кінцевого користувача. До 2011 року існували так звані RSS-оповідачі, які об'єднували лише текстові стрічки новин та перетворювали їх на аудіозаписи для прослуховування в режимі офлайн. Синдигований вміст, який збирає та інтерпретує агрегатор, зазвичай подається у формі RSS або інших даних у форматі XML, таких як RDF / XML або Atom.

1.2 Поняття та історія створення технології автоматизації бізнес-процесів компанії із використанням технології RPA

RPA – Robotic Process Automation. RPA – це форма технології автоматизації бізнес-процесів, заснована на понятті ПЗ ботів. У традиційних системах автоматизації розробник програмного забезпечення створює певний алгоритм для автоматизації завдання або взаємодій з внутрішньою системою з використанням внутрішніх інтерфейсів прикладного програмування (API). RPA у цьому відрізняється т.я. займається тим, що емулює дії користувача в інтерфейсі програми (GUI) за засобами спостереження за роботою користувача та записуючи порядок дій, які він здійснив. Даний спосіб автоматизації дозволяє

взаємодіяти з продуктами, якими раніше не було можливості це зробити з технічних причин: відсутність API, закритий код, застаріла програма і т.я. Інструменти RPA за останні кілька років зробили величезний крок у своєму розвитку, і якщо раніше це було щось слабо застосовне і створене виключно для специфічних завдань, то зараз ринок пропонує досить широкий вибір інструментів, які мають великий функціонал для роботи в багатьох напрямках, нітрохи не поступаючись будь-якій іншій системі.

RPA пропонує два види інтеграції із середовищем запуску процесу.

Допоміжна автоматизація, де програмне забезпечення RPA автоматизує інші програми, що працюють на настільному комп'ютері співробітника у тісній співпраці зі співробітником оскільки саме користувач вирішує, коли слід залучити робота для тієї чи іншої роботи, такі роботи називаються роботами-помічниками. Другий вид - це повна автономність робота від співробітника та його робочої станції оскільки виконання робіт відбуватиметься на окремому сервері і не потребує взаємодії з користувачем. Кожна з представлених моделей має свої переваги та обмеження.

Поетапне програмування дій бота, що піддається алгоритмізації або стандартизації діям користувача допомагає надати програмному застосунку можливість працювати з тими ж самими графічними інтерфейсами програм та додатків, з якими користувач здійснює взаємодію на своєму робочому місці, але робить це з більшою швидкістю та у більшому обсязі (на порядок швидше за людину, 24 години на добу 7 днів на тиждень. RPA дозволяє інтегрувати у свій процес використання несумісних на програмному рівні інформаційних систем, обробляючи дані кожної з них та використовуючи дані однієї системи в іншій системі (наприклад, автоматизувати процес заповнення електронних форм однієї інформаційної системи на основі Microsoft Excel, використовуючи дані, доступні у форматі PDF в інтерфейсі іншої інформаційної системи). Згідно з прогнозами аналітичної компанії Gartner, світові темпи роботизації робочих процесів RPA (Robotic process automation) в найближчі роки будуть невпинно зростати. 13

листопада 2018 року Gartner опублікувала результати дослідження ринку програмного забезпечення для роботизованою автоматизації процесів (RPA). Згідно з попередніми даними, в 2018 році світові витрати на такі програмні застосунки склали \$ 680 млн, збільшившись на 58% щодо попереднього року. «Працюючи з кінцевими користувачами організації впроваджують технологію RPA в якості швидкого і легкого засоби автоматизації завдань, що вимагають ручної праці, - говорить віце-президент Gartner Кеті Торнбом (Cathy Tornbohm).

«Деякі співробітники будуть продовжувати виконувати рутинні завдання, які вимагають від них ручної вирізки, вставки і зміни даних. Але, коли RPA-інструменти виконують ці дії, межа помилок скорочується, а якість даних збільшується.»

Кеті Торнб віце президент Gartner

Згідно з прогнозами компанії Forrester, до 2021 року ринок RPA може вирости до \$ 3 млрд. Основними факторами зростання експерти Forrester називають розвиток RPA-систем і їх адаптацію під більшість корпоративних процесів, а також синергію RPA-систем і штучного інтелекту, що дозволяє впливати на процеси прийняття управлінських рішень.

1.3 Технологія Web Scraping та її різновиди для впровадження системи пошуку резюме за даними із відкритих архівів

Веб-скрейпінг — це процес автоматичного отримання даних або збору інформації із мережі інтернет або з окремих ресурсів. Це сфера з активними розробками, які мають спільну мету з баченням семантичної мережі, амбітну ініціативу, яка все ще потребує проривів в обробці тексту, семантичному розумінні, штучному інтелекті та взаємодії людини та комп'ютера.

Перший приклад роботи сканування веб-сторінок датується 1993 роком, який був визначним роком для цієї технології. У червні того ж року Метью Грей розробив World Wide Web Wanderer Offsite Link для вимірювання розміру

Інтернету. Пізніше того ж року це було використано для створення індексу під назвою «Wandex», і це дозволило створити першу веб-пошукову систему. Сьогодні ми сприймаємо це як належне з великими пошуковими системами, які майже миттєво надають безліч результатів. Примітно, що до того, як була запущена технологія веб-скрейпінгу JumpStation, збір даних здійснював адміністратор вручну, який збирав і форматував набори даних, відповідали б тому, що шукали користувачі.

Мануальне копіювання та вставлення

Найпростіша форма веб-скрейпінгу – це та, в якій людина вручну копіює та вставляє дані з веб-сторінки у текстовий файл або електронну таблицю. Іноді навіть найкраща технологія веб-скрейпінгу не може замінити мануальну обробку, але іноді це може бути єдиним працездатним рішенням, коли наприклад веб-сайти для веб-скрейпінгу встановлюють програмні запобіжники для перешкоджання автоматизації процесу отримання інформації.

Відповідність шаблону тексту

Простий, але потужний підхід до вилучення інформації з веб-сторінок може бути заснований на команді UNIX `grep` або засобах узгодження регулярних виразів мов програмування (наприклад, Perl або Python).

HTTP програмування

Розмітку статичних або динамічних веб-сторінок можна отримати, відправивши HTTP-запити на віддалений веб-сервер за допомогою програмування сокетів.

Розбір HTML

Багато веб-сайтів мають великі колекції сторінок, які динамічно генеруються з базового структурованого джерела, наприклад бази даних. Дані однієї категорії зазвичай кодуються на схожих сторінках за допомогою загального сценарію або шаблону. У аналізі даних програма, яка виявляє такі шаблони в певному джерелі інформації, витягує його вміст і переводить його у реляційну форму, називається

обгорткою. Алгоритми генерації обгортки припускають, що вхідні сторінки системи індукційної обгортки відповідають загальному шаблону і що їх можна легко ідентифікувати за допомогою загальної схеми URL. Крім того, деякі напівструктуровані мови запитів даних, такі як XQuery та HTQL, можна використовувати для розбору HTML-сторінок, а також для отримання й перетворення вмісту сторінки.

Розпізнавання семантичних анотацій

Сторінки, які було оброблено технологією веб-скрейпінгу, можуть містити метадані або семантичні розмітки та анотації, які можна використовувати для пошуку певних фрагментів даних. Якщо анотації вбудовані в сторінки, як це робить ПЗ Microformat, цей прийом можна розглядати як окремий випадок розбору DOM. В іншому випадку анотації, організовані в семантичний шар, зберігаються та керуються окремо від веб-сторінок, тому ПЗ для веб-скрейпінгу можуть отримати схему даних та інструкції з цього шару перед тим, як вибирати сторінки.

Аналіз веб-сторінки із використанням комп'ютерного зору

Існують ПЗ з використанням машинного навчання та комп'ютерного зору, які намагаються ідентифікувати та витягти інформацію з веб-сторінок, інтерпретуючи сторінки візуально, як це може зробити людина.

Як працює веб-скрейпінг?

По-перше, веб-скрейперу буде надано одну або кілька URL-адрес для завантаження перед розподіленням інформації. Потім скрепер завантажує весь HTML-код відповідної сторінки. Більш просунуті скребки відобразатимуть весь веб-сайт, включаючи елементи CSS та Javascript.

Потім скрепер отримає всі дані на сторінці або конкретні дані, вибрані користувачем перед запуском проекту.

В ідеалі користувач пройде процес вибору конкретних даних, які йому потрібні, зі сторінки. Наприклад, ви можете отримати інформацію про ціни та моделі на сторінці продукту Amazon, але не обов'язково вас цікавлять огляди продуктів. Невдовзі, веб-скрепер виведе всі зібрані дані у формат, який є більш корисним для користувача.

Більшість веб-скребків виводить дані в електронну таблицю CSV або Excel, тоді як більш просунуті скрепери підтримують інші формати, такі як JSON, які можна використовувати для API.

1.4 Історія створення програмних рішень систем фільтрації, агрегації та репрезентації резюме

Першими великими, професійними сайтами з пошуку роботи прийнято вважати Monster.com або CareerBuilder.com, який складався з двох полів, структурно подібних до газетних оголошень, для роботодавців і для шукачів. Сайти служили як інформаційне поле для пошуку контактів і в них була відсутня будь-яка форма авторизації для ідентифікації претендента або створення статичного профілю компанії. Оскільки дана структура сайту була застарілою як з технічного, так і з візуального боку, самостійні компанії прийняли стратегічне рішення про створення своїх продуктів, які б нівелювали недоліки вищезгаданих сайтів.

Наведемо приклад одного з перших сайтів з пошуку роботи на території України - roboota.ua.

Сайт roboota.ua був створений групою українських програмістів у 2000 році.

У 2006 році стратегічним партнером та інвестором roboota.ua стає польська компанія Grupa Pracuj Ltd, лідер інтернет-рекрутингу Центральної Європи. Цього ж року roboota.ua стає партнером міжнародної рекрутингової мережі The Network, представленої у понад ста країнах.

Протягом 2007-2014 років roboota.ua провела низку спеціалізованих подій для HR-фахівців та партнерів із міжнародними експертами.

У 2017 році robot.ua, крім Києва, відкрила представництва у Дніпрі, Харкові, Чернігові та інших містах. Блог для шукачів став окремим інтернет-виданням «The Point», яке спрямоване на професійне та кар'єрне розвиток шукачів.

На даний момент вищенаведеним сайтом користуються 58% кандидатів під час пошуку роботи в Україні, що є абсолютним лідерством на поточному ринку даних додатків.

З іншого боку, можна навести приклад продукту, який на даний момент є найбільшою у світі професійною соціальною мережею - LinkedIn.

У грудні 2002 року була створена ділова соціальна мережа LinkedIn, сайт якої було спроектовано та відкрито доступ усім користувачам мережі інтернет у травні 2003 року.

Головною ідейною концепцією проекту було створення ресурсу, на якому можна було б розмістити своє резюме, знайти своїх колег, попросити у них рекомендацію. Основою LinkedIn була проєктована ідея спілкування професіоналів, колег по роботі. Генеральний директор компанії вважав, що створення та підтримка ділових зв'язків — це важливий елемент життя та діяльності будь-якого спеціаліста, що більшості людей, орієнтованих на кар'єрне зростання або просто зацікавлених у пошуку хорошої роботи, важливо зберегти можливість зв'язатися зі своїми колишніми колегами, наприклад, для того, щоб попросити рекомендацію.

Сервіс став швидко затребуваний завдяки професійному аналізу тогочасної ситуації на ринку праці та виявленим ґрунтовним чинникам, один із них — американські спеціалісти, для кар'єрного зростання активно змінювали місце роботи оскільки, вертикальне зростання в одній компанії не вимагало публічності у претендента. У той же час, для переходу з однієї компанії в іншу від претендента були потрібні і рекомендації, і публічний профайл, друге було критично важливим для можливості бути знайденим рекрутерами іншої компанії.

Через місяць після запуску сайту на його сторінках зареєструвалося 4,5 тис. людей. Спочатку зростання проекту було досить повільним - наприклад, до 2004 року на сторінках LinkedIn зареєструвалося 1 млн користувачів. Далі користувальницька база зростала на 2-3млн щорічно.

До 2008 року аудиторія LinkedIn налічувала 20 млн людей. На сайті стали реєструватися не тільки IT-фахівці, а й люди, які працюють в інших галузях діяльності. А рекрутери зацікавилися можливостями доступу до баз даних людей, зареєстрованих на LinkedIn.

Водночас стало зрозуміло, що акцент сервісу на професіоналів та використання проекту для утилітарних цілей дозволить LinkedIn уникнути проблем, з якими зіткнулися соціальні мережі MySpace чи Facebook. Ці два проекти, орієнтовані на молодих користувачів, довго перебували в пошуках інструментів монетизації, а MySpace — це зовсім приклад соціальної мережі, яка швидко завоювала аудиторію, але так само її втратила.

1.5 Міжнародний досвід на прикладі систем агрегації резюме

Для підкреслення актуальності розроблюваного програмного застосунку нижче наведено приклади існуючих програмних рішень, якими користуються закордонні представництва та міжнародні компанії для пошуку висококваліфікованих працівників, використовуючи принцип автоматизації підбору працівників.

Приклад 1

roboota.ua — веб-сайт, який представляє Україну в мережі провідних світових сайтів The Network за такими критеріями, як кількість і якість вмісту. Щомісячна відвідуваність перевищує чотири млн. унікальних відвідувачів, а загальна складає понад 10 млн.

Сайт roboota.ua було створено програмістів із України у 2000 році.

У 2006 році стратегічним партнером та інвестором roboota.ua стала польська компанія Grupa Pracuj Ltd, яка є міжнародно визнаним лідером із інтернет-рекрутингу у регіоні Центральної Європи.

Протягом 2007—2014 років robota.ua провела ряд спеціалізованих подій для HR-фахівців та партнерів із міжнародними експертами.

У 2017 році robota.ua, окрім Києва, відкрила представництва у Дніпрі, Харкові, Чернігові та інших містах. Веб-блог для шукачів став окремим інтернет-виданням «The Point», яке спрямоване на професійний та кар'єрний розвиток шукачів.

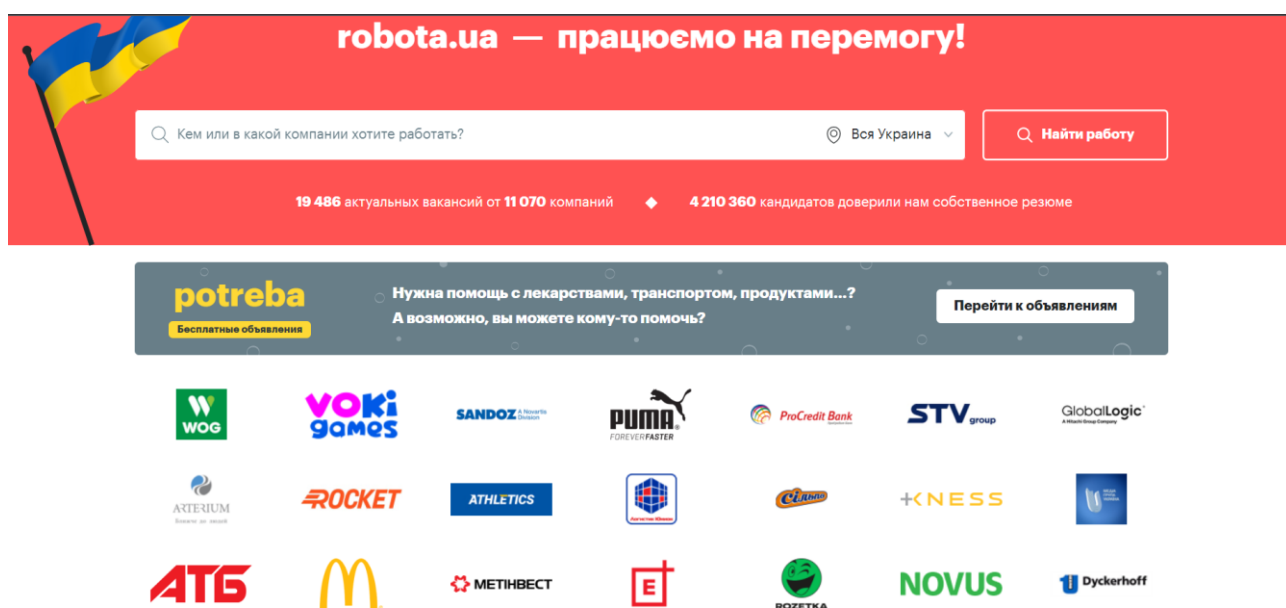


Рисунок 1.1 – Інтернет-сайт robota.ua

Сайт robota.ua має такі переваги:

- великий попит зі сторони українських роботодавців;
- широкий спектр стандартизованих даних для задання певних критеріїв резюме кандидата, що допомагає отримувати інформацію пошуку швидше аніж у неструктурованій системі;
- простота використання інтерфейсу ресурсу;
- висока репутація за кордоном;
- часті акції, знижені ціни для виставлення вакансій роботодавцями;

Приклад 2

Приклад українського продукту із пошуку та агрегації резюме та вакансій кандидатів – jooble.org

Привабливість цього сайту в тому, що він представляє собою сайт-агрегатор багатьох міжнаціональних централізованих сайтів пошуку вакансій.

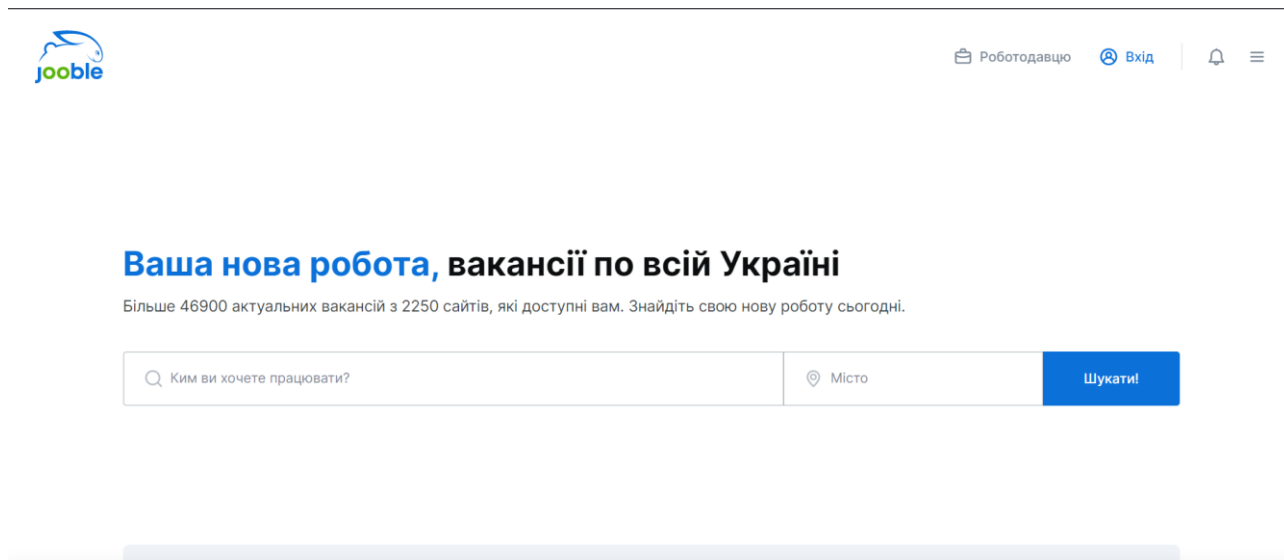


Рисунок 1.2 – Сайт-агрегатор пошуку роботи – Jooble

Jooble був заснований в 2006 році командою студентів із України: Романом Прокоф'євим і Євгеном Собакар'євим. В 2008 програмний сервіс працював у Росії, Казахстані та Білорусі. Станом на 2010, сервісом можна було скористатись у 24 країнах. На даний момент Jooble доступний в 71 країні. За 2014 рік сервісом скористались понад 800 млн користувачів.

Jooble знаходиться у ТОП-1000 найвідвідуваніших сайтів світу, а з грудня 2019 є другим найвідвідуванішим сайтом з пошуку роботи з 90 млн відвідувачів на місяць в середньому та більш ніж 950 млн відвідувачів протягом всього 2019 року.

Jooble збирає оголошення про вакансії з інших централізованих агрегаторів сторінок компаній, газет, рекрутингових агенств. Користувач розпочинає пошук на сайті з введення міста і назви вакансії. За допомогою фільтрів можна сортувати оголошення за віддаленістю від міста, датою, зарплатою і назвою

підприємства-роботодавця. Алгоритм пошук оголошень про роботу враховує сотні параметрів і підбирає для користувача релевантні і актуальні вакансії. Jooble пропонує вакансії як з повною зайнятістю, так і з частковою, а також практики і фрілансерські вакансії. Клікнувши на оголошення, користувач перенаправляється на відповідну сторінку компанії, де він зможе знайти контакти роботодавця і подати резюме.

Приклад 3

Sniper AI використовує машинне навчання, щоб відбирати кандидатів і зіставляти їх резюме зі специфікаціями вакансії із конкурентною швидкістю та 95%+ точністю. За підтримки Recruitment SMART, британського технічного стартапу HR, Sniper AI можна легко інтегрувати із системами відстеження кандидатів для легкої перевірки резюме. Sniper AI забезпечує 53% внутрішнього скорочення робочої сили, що дозволяє рекрутерам витратити менше часу на перевірку резюме. З такими клієнтами, як Infosys, Vodafone, Capgemini тощо, цей інструмент є досить відомим серед індустрії.

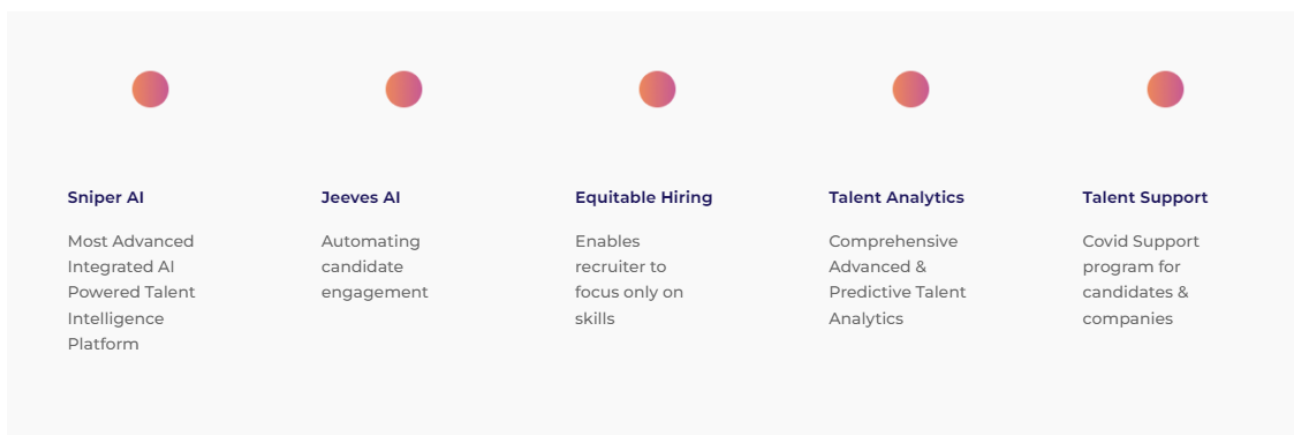


Рисунок 1.3 – Основні можливості програмного застосунку. (Оригінал тексту взято із веб-сторінки інструменту)

РОЗДІЛ 2. АНАЛІЗ ПІДХОДІВ І ПРИНЦИПІВ ПОБУДОВИ ПРИКЛАДНОГО ЗАСТОСУНКУ ПОШУКУ РЕЗЮМЕ

2.1 Аналіз принципів побудов застосунків автоматизації бізнес-процесів, використовуючи технологію RPA

Перед проектуванням автоматизованої інформаційної системи слід розглянути системи, які мають у своєму функціоналі вирішення подібних завдань. Таким чином, мова піде про HRM-системи. HRM-системи – це автоматизовані системи управління персоналом, призначені як вирішення завдань стандартних кадрових операцій, але й забезпечують роботу з якісними показниками персоналу.

Основні завдання HRM-систем:

1. Автоматизація процесів, пов'язаних з обліком персоналу, і навіть розрахунки.
2. Виявлення та збереження цінних, а також залучення нових співробітників.

Були розглянуті такі HRM-системи:

- SAP ERP HCM;
- Oracle HRMS.

Перераховані вище системи мають функцію управління персоналом. Проте загальною рисою всіх цих систем є адаптація для роботи в організації, тобто вони включають обліково-розрахункові функції: розрахунок заробітної плати, організація розкладу і т. д. Наявність цих функцій ускладнює впровадження даних систем у відділ пошуку спеціалістів компанії, головною та єдиною метою якої є автоматизоване зіставлення вакансій та анкет здобувачів.

Платформа UiPath заснована на локальному та хмарному методі розгортання, і завдяки цьому можливо розгортати і керувати всім в одному конкретному місці, що полегшує управління автоматизованими бізнес-процесами і процесами віртуальної робочої сили. Він також має широкий діапазон продуктивності, оскільки UiPath забезпечує розширюваність і стандартизованість, що дає новий рівень автоматизації, який використовується для мінімальної вартості

експлуатації, а також забезпечує кращу турботу про використання ІТ ресурсів. Якщо для різних бізнес-підрозділів існують різні користувачі, з UiPath можна створити один сервер, який буде захищеним, приватним і автономним, що забезпечує високий рівень безпеки.

Автоматизація процесів в UiPath - це створення сценаріїв роботизації шляхом перетягуванням необхідних дій з бібліотеки (панелі інструментів) в workflow (своєрідний модуль, що виконується при запуску), як це робиться в інших візуальних інструментах. Також, UiPath надає можливість записувати дії користувача, після чого автоматично переводить цю запис в елементи відповідних дій з бібліотеки. UiPath підтримує більшість видів додатків: звичайні десктопні програми MS Windows, WEB сайти, Java додатки, віртуальні середовища (в тому числі Citrix), а також "зелені екрани" мейнфреймів і iSeries. У UiPath також присутні вбудовані засоби розпізнавання символів на зображенні (OCR), які часто необхідні при роботі з документами або віртуальними середовищами. Стандартний програмний пакет надає кошти OCR технології Google і Microsoft. Але при необхідності, додатково можна придбати ліцензію для технології Abbyy .

2.2 Технології RPA та DataScraping як основне архітектурне рішення для побудови застосунку

Переваги технології RPA:

- RPA-процес має можливість взаємодії з будь-якою програмною системою, у той час як програмні рішення фільтрації, агрегації та репрезентації новин, описані у роботі, мають можливість взаємодіяти лише із обмеженою кількістю програмних систем.
- Інструменти технології RPA мають можливості інтеграції з програмною технологією OCR, тому ви можете виконувати процеси на віддаленому робочому столі або в середовищі Citrix.
- Кардинальна відмінність, існуючих програмних систем, описаних у роботі, та технології RPA полягає в тому, що в завдання традиційної

автоматизації входить back-end діяльність. Навпаки, роботизована автоматизація бізнес-процесів програмно взаємодіє тільки з інтерфейсом користувача.

- Використання технології RPA і Data Scraping дозволяє взаємодіяти з сайтами без використання програмного API або REST-запитів, що відкриває допоміжні можливості для автоматизації людської праці.

- При побудові програмної системи, яка ґрунтується на використанні технології RPA, розробник апілює ширшим функціоналом взаємодії із середовищем операційного пристрою.

- Програмна система, побудована на технології, має можливість запуску декількох екземплярів програмного процесу на декількох операційних пристроях.

Для отримання цільової інформації із веб-документів HTML, XHTML, XML стандартними засобами PHP найчастіше застосовуються програмні інтерфейси (API) DOM (Document Object Model), SAX (Simple API for XML), XMLReader, SimpleXML. Ці типи інтерфейсів включені стандартно в дистрибутиви PHP, базуються на бібліотеці libxml2 проекту GNOME. За способом завантаження веб-документа API бувають або з завантаженням в пам'ять цілком і побудовою об'єктної структури, або з потоковою завантаженням (послідовно по частинах), в останньому випадку розрізняють передає (push) аналізатор (на основі подій) і приймає (pull) аналізатор (курсорний).

Data Scraping (парсинг інформаційної сторінки) можна здійснити кількома способами, залежно від того, для чого використовується процес. Наприклад, за допомогою Java людина може копіювати та вставляти вихідний код з однієї програми у свою власну, якщо вона має шлях прямого доступу до неї.

Data scraping (парсинг інформаційної сторінки) дозволяє користувачеві витягувати дані відображення на екрані з певного елемента інтерфейсу або документів. Для отримання всього тексту на сторінці, неформатованого, або тексту на сторінці, відформатованого з точним позиціонуванням, можна використовувати різні методи. Програмні засоби, що здійснюють можливість зчитування інформації можуть базуватися на таких додатках, як Selenium або

PhantomJS, що дозволяє користувачам отримувати інформацію з HTML у браузері.

2.3 Різні види забезпечення побудови прикладного застосунку із пошуку резюме за джерелом відкритих даних

Програмне забезпечення повинно в першу чергу виконувати головну поставлену перед ним задачу. В дипломній роботі перед прикладною програмою із пошуку резюме кандидатів за критеріями стоїть задача надати користувачу системи задати критерії пошуку у зручному для нього вигляді, та отримати результати пошуку за доволі короткий час.

До основних функцій програмного забезпечення відносяться:

- службові функції виконують непрямі, неявні функції програмного забезпечення (наприклад, перевірка коректності введених даних, зберігання результатів попередніх запитів користувача);

- основні функції управління і обробки даних виконують прямі функції: введення первинної інформації(критеріїв пошуку), пошуку інформації за джерелами відкритих даних, логічний аналіз, видача підсумкових даних.

Розробка розрахована не тільки на вузьке коло споживачів, а й на широке коло клієнтів. Прикладний застосунок із пошуку резюме за користувацькими критеріями може використовуватися у напрямках діяльності пошуку кадрів, як у масштабних компаніях так і у приватному використанні фізичними особами.

До споживчих характеристик продукту відносяться:

- простота у використанні;
- зручність пошуку;
- швидка обробка запитів користувачів;
- ефективність.

Технічне забезпечення є комплексом технічних засобів, що призначені для роботи інформаційних систем, а також має відповідну документацію на представлені засоби та технологічні процеси.

Конфігурація інформаційних систем, що проектуються і обладнання, яке використовується, повинні забезпечити можливість нарощування за рахунок розширення апаратної і програмної частин без порушення їх працездатності.

Оптимальна конфігурація комп'ютера для розробки прикладного застосунку представлена у таблиці 2.1.

Таблиця 2.1 – Конфігурація комп'ютера

Параметр	Значення
Корпус	Корпус GameMax MT-306-NP-U3-CR
Процесор	Процесор Intel Pentium Gold G6405 4.1 GHz / 4 MB s1200 BOX
Кулер для процесору	Socket1200 Кулер ID-Cooling SE-224-XT Black
Вентилятор	Be quiet! Pure Wings 2 140 mm PWM high-speed
Материнська плата	Socket1200 Материнська плата Asus Prime H510M-E (s1200, Intel H510, PCI-Ex16)
Модуль оперативної пам'яті	4ГБ DDR4 SDRAM Kingston „ValueRAM” KVR800D2N6/2G

Продовження таблиці 2.1

Жорсткий диск	500ГБ Seagate „Barracuda 7200.12 ST3320418AS” 7200об./мин., 64МБ (SATA III)
Пристрій читання карт пам'яті	CF/MD/SM/MMC/SD/MS/MS Pro Sony „MRW620”, в 3.5” відсік, чорний (USB2.0)
Оперативна пам'ять	8 Gb
Монітор	Asus 24”
Мережа інтернет	Ethernet

При виборі технічних засобів необхідно враховувати наступні критерії:

- сумісність технічних засобів з операційною системою;
- обсяги інформації, що обробляється, їх структура, час обробки;
- необхідність наявності засобів комунікації;
- структура завдань, що вирішуються, їх взаємозв'язок;
- вимоги до умов експлуатації технічних засобів;
- система повинна функціонувати цілодобово і безперервно;
- продуктивність, параметри і характеристика технічних засобів.

РОЗДІЛ 3. ПРОЕКТУВАННЯ, РОЗРОБЛЕННЯ ТА ВПРОВАДЖЕННЯ ПРОГРАМНОЇ СИСТЕМИ ПОШУКУ РЕЗЮМЕ

3.1 Опис концепції та програмної реалізації технології Data Scraping при розробці прикладної системи

Процес Data Scraping веб-сторінки передбачає структуроване вилучення даних із HTML-коду. Scraping - це завантаження сторінки (що робить браузер, коли користувач переглядає сторінку). Тому сканування веб-сторінок є головним компонентом технології для отримання сторінок для подальшої обробки.

Вміст сторінки може бути проаналізований, здійснений пошук, переформатування, копіювання в електронну таблицю або завантаження в базу даних. Веб-зчитувачі, як правило, виймають щось зі сторінки, щоб використати це з іншою метою в окремому програмному модулі.

Як було написано раніше, технологія виробляє зчитування і структурування даних аналізуючи html-структуру сайту для виявлення патернів, які визначають та диференціюють, в нашому випадку, новинні заголовки.

Приклад роботи технології описано нижче:

Для позначення даних сторінки і подальшим виділенням патернів за допомогою технології Data Scraping, використовуються теги мови розмітки HTML, представлені у деревоподібній, об'єктній формі.

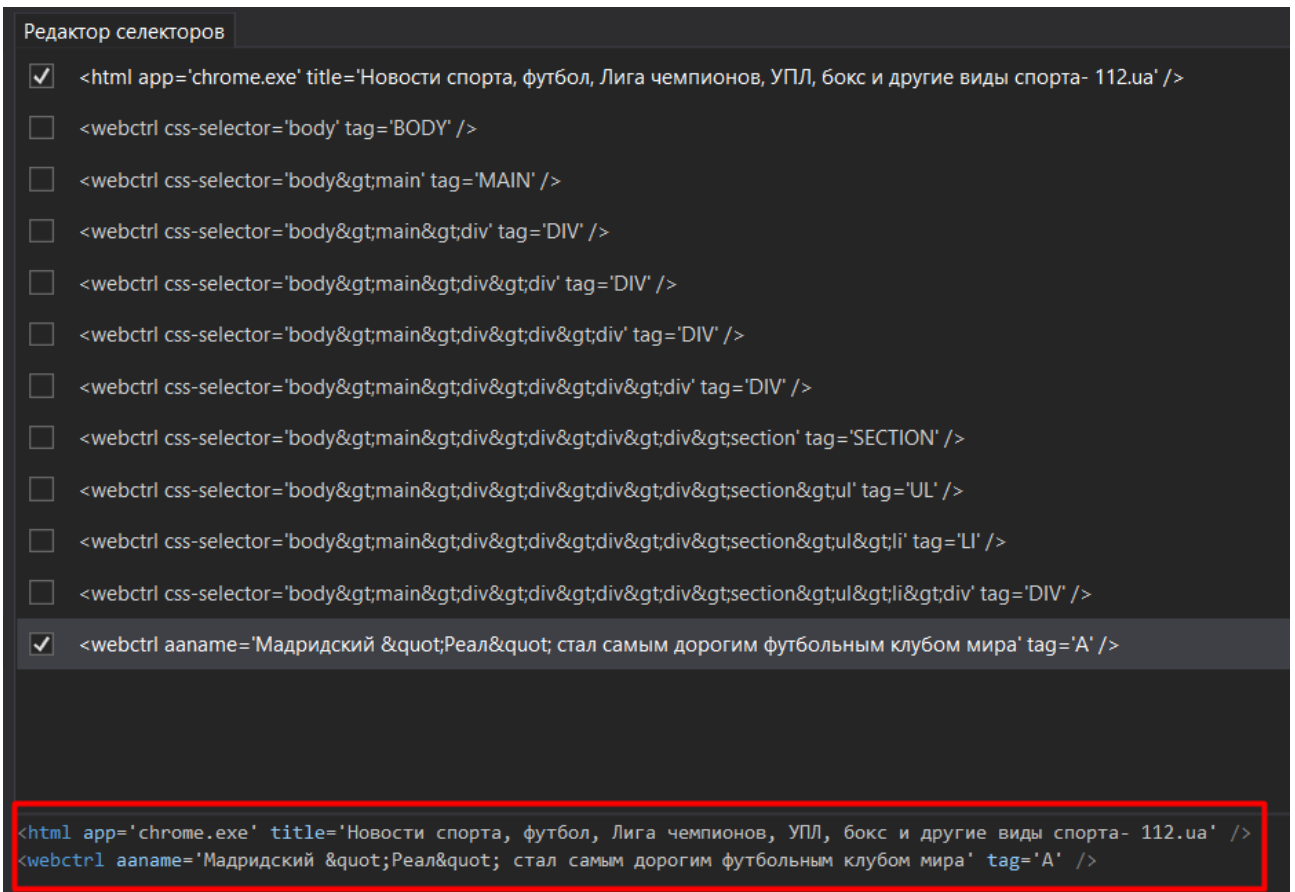


Рисунок 3.1 - Приклад отриманої інформації із коду HTML – сторінки, який диференціює кожний об’єкт сторінки для програмного середовища Data Scraping

Для того щоб виявити патерн на сторінці архівних новин, був проведений аналіз HTML-тегів сторінки і приклади послідовних записів в списку новин були подані входить аргументами в програмне забезпечення UiPath, яке, в свою чергу, визначило патерн записів на сторінці, які відповідають, наведеним на вхід , прикладів.

Нижче на малюнках 3.2 та 3.3 відображено визначення вступних даних для системи UiPath і подальше визначення патерну системою.

Виділені записи новин - це обумовлені прикладною програмною системою дані для проведення парсинга у структурований формат формат DataTable.

АРХИВ

Мадридский "Реал" стал самым дорогим футбольным клубом мира

08:32 18.05.2021

Сергей Ребров может возглавить греческий АЕК

16:53 17.05.2021

В игре против Бахрейна посмотрим, на каком уровне сборная Украины под нагрузками, - Шевченко

10:22 17.05.2021

"Лестер" впервые в своей истории выиграл Кубок Англии

21:13 15.05.2021

Гол и ассист Малиновского помогли "Аталанте" выйти в Лигу чемпионов

19:18 15.05.2021

"Челси" - "Лестер": Онлайн-трансляция финала Кубка Англии

19:16 15.05.2021

Элина Свитолина проиграла Швентек в четвертьфинале турнира WTA в Риме

14:40 15.05.2021

"Челси" - "Лестер": Сегодня состоится финал Кубка Англии

10:31 15.05.2021

Лидеры группы U2 и диджей Мартин Гаррикс представили клип на официальную песню Евро-2020

12:27 14.05.2021

Владимир Шаран покинул "Александрию"

Рисунок 3.2 - Пример виділення першочергового елемента для уточнення патерну заголовків новин на структурованій веб-сторінці

АРХИВ

Мадридський "Реал" став самим дорогим футбольним клубом мира

08:32 18.05.2021

Сергей Ребров может возглавить греческий АЕК

16:53 17.05.2021

В игре против Бахрейна посмотрим, на каком уровне сборная Украины под нагрузками, - Шевченко

10:22 17.05.2021

"Лестер" впервые в своей истории выиграл Кубок Англии

21:13 15.05.2021

Гол и ассист Малиновского помогли "Аталанте" выйти в Лигу чемпионов

19:18 15.05.2021

"Челси" - "Лестер": Онлайн-трансляция финала Кубка Англии

19:16 15.05.2021

Элина Свитолина проиграла Швентек в четвертьфинале турнира WTA в Риме

14:40 15.05.2021

"Челси" - "Лестер": Сегодня состоится финал Кубка Англии

10:31 15.05.2021

Лидеры группы U2 и диджей Мартин Гаррикс представили клип на официальную песню Евро-2020

12:27 14.05.2021

Владимир Шаран покинул "Александрию"

08:44 14.05.2021

Рисунок 3.3 - Приклад графічного відображення створеного програмного-патерну для отримання заголовків новин на структурованій веб-сторінці

Після визначення патерну (Рисунок 3.3) - дані можуть бути вивантажені в формат даних мови програмування C# - DataTable.

DataTable є центральним об'єктом у бібліотеці ADO.NET. Інші об'єкти, які використовують DataTable, включають DataSet та DataView.

Після створеного програмного патерну, використовуючи програмний інтерфейс оболонки для розробки прикладних програмних модулів UiPath Studio, що ґрунтується на технології RPA, можна відстежити коректність отриманих новинних заголовків, а також побачити програмне відображення написаного патерну в форматі XML-документу.

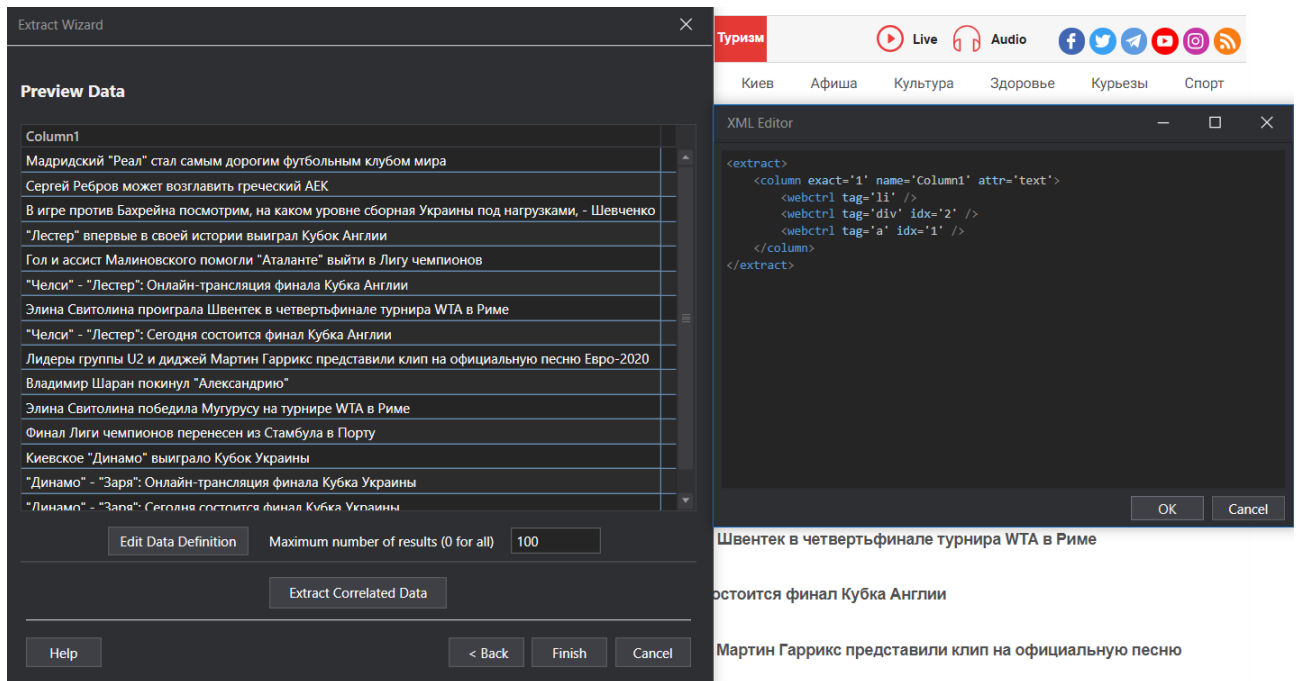


Рисунок 3.4 - Приклад фінальної фази проведення Data Scraping із веб-сторінки – відображення отриманої інформації у UI-інтерфейсі оболонки UiPath Studio.

3.2 Технічне завдання та постановка задачі програми пошуку та структурування даних резюме із відкритих джерел

Головна мета дипломного проекту є розробка програмного забезпечення для отримання та структурування даних резюме з відкритих джерел, за корисувачькими критеріями .

Вимоги до функціональних характеристик.

Програмний продукт для пошуку та структурування даних резюме із відкритих джерел повинен забезпечувати можливість виконання наведених нижче функцій:

- отримання переліку резюме кандидатів із усіх, заданих користувачем джерел;
- зміна критеріїв пошуку кандидатів у зручному для користувача форматі;
- можливість переглянути резюме у базі архіву, з якого воно було отримано програмою.

Вимоги до користувача.

Для експлуатації інтерфейсу системи керування програмою від користувача не повинно вимагатися спеціальних технічних навичок, знання технологій або програмних продуктів, за винятком загальних навичок роботи з персональним комп'ютером та стандартом роботи із продуктом MS Office(Excel).

Вимоги до розмежування доступу.

Інформація, що була отримана із відкритих джерел – зберігається лише на комп'ютері користувача, який взаємодіє із програмою тому в останньої немає додаткових переваг щодо розмежування відповідальності та конфіденційності окрім стандартів безпеки користування персональним комп'ютером користувача.

Користувачі мають доступ тільки до загальнодоступної інформації.

Користувач може: додавати або видаляти критерії пошуку резюме. Доступ до технічної частини має здійснюватися з використанням стандартів взаємодії із програмним комплексом рішень UiPath Studio.

Вимоги до інтерфейсу користувача.

Користувач повинен мати можливість:

- отримання переліку резюме кандидатів із усіх, заданих користувачем джерел;
- зміни критеріїв пошуку кандидатів у зручному для користувача форматі;
- можливість переглянути резюме у базі архіву, з якого воно було отримано програмою.

Стадії і етапи розробки.

Розробка має бути проведена в три стадії:

- розробка технічного завдання;
- розробка робочого проекту;
- впровадження.

Етапи розробки:

На стадії розробки технічного завдання має бути виконаний етап обґрунтування необхідності розробки програми, етап науково-дослідної роботи та етап розробки та затвердження технічного завдання.

На стадії робочого проектування мають бути виконані перелічені нижче етапи робіт:

- розробка програми пошуку та структурування резюме із відкритих джерел;
- розробка документації;
- тестування програмного застосунку.

На стадії впровадження має бути виконаний етап підготовки та передачі програми.

Для передачі програмного застосунку мають бути надані наступні документи:

- технічне завдання;
- пояснювальна записка.

Порядок проведення тестування:

- введення критеріїв пошуку резюме у таблицю;
- запуск програми пошуку;
- оцінка правильності виконання.

3.3 Опис і склад програмних модулів програмної системи автоматизації пошуку резюме за визначеними користувацькими критеріями.

Програма пошуку резюме за заданими критеріями користувача представляє собою прямий процес, основні етапи якого описуються нижче:

Зміна критеріїв пошуку кандидатів.

Користувач заходить у середу MS Office, відкриває таблицю із заздалегідь заданою програмою структурою та заповнює стільки рядків, скільки вакансій потрібно закрити. При незаповненому рядку – процес розпізнає його як неважливий та продовжить пошук із усіма можливими типами значень цього поля.

Можливість додавати архіви для пошуку інформації

За заздалегідь написаною інструкцією, користувач має можливість взаємодіяти із low-code середовищем, що дозволяє затрачувати менше ресурсів для того щоб змінювати процес, тому що low-code структура має за ідею спрощення процесів класичного програмування.

Можливість переглянути резюме у базі архіву, з якого воно було отримано програмою

Після запуску програми, із введеними заздалегідь користувацькими даними, процес продивляється записані у логіку програми сервіси, та виділяє найбільш важливі та близькі за критеріями резюме та виводить список посилань на них у таблиці користувача. Тому користувач, може переглянути детально кожного кандидата, та приймати рішення щодо подальшої взаємодії із ним.

Перелік архівів відкритих даних:

- jooble
- rabota.ua
- work.ua

Архітектура розробленого програмного застосунку

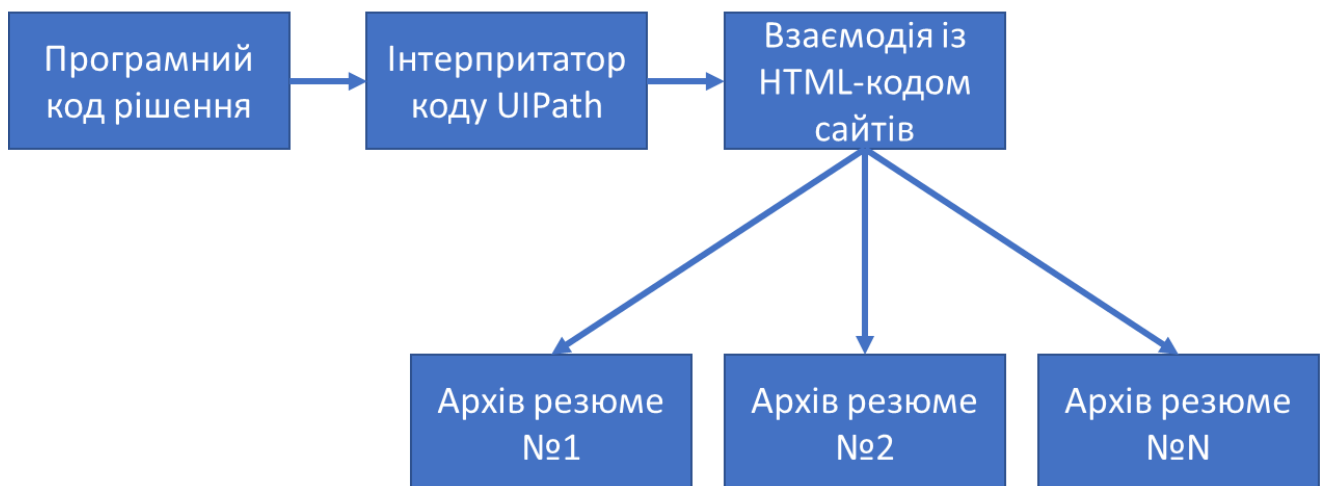


Рисунок 3.5 - Архітектура побудованого програмного застосунку на платформі UiPath

Програмний код рішення розроблюється у спеціальному середовищі для розробки RPA-рішень UiPath, яке має можливість створювати програмний продукт, використовуючи low-code підхід. Код рішення зберігається у файлах типу .xml.

Після запуску розробленого програмного коду(використовується функція продукту UiPath) – активується інтерпретатор коду який перетворює описану логіку у графічному редакторі середовища у машинний код, який використовується комп'ютером для виконання вказаних операцій.

Взаємодія із HTML-кодом сайтів відбувається використовуючи елементи коду як точки зв'язку, ґрунтуючись на чому програма може виконувати такі дії, як натискання на елементи інтерфейсу курсором миші, або введення тексту у поля.

Для позначення даних сторінки і подальшим виділенням патернів за допомогою технології Data Scraping, використовуються теги мови розмітки HTML, представлені у деревоподібній, об'єктній формі.

```

Редактор селекторов
 <html app='chrome.exe' title='Новости спорта, футбол, Лига чемпионов, УПЛ, бокс и другие виды спорта- 112.ua' />
 <webctrl css-selector='body' tag='BODY' />
 <webctrl css-selector='body&gt;main' tag='MAIN' />
 <webctrl css-selector='body&gt;main&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl css-selector='body&gt;main&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div&gt;div' tag='DIV' />
 <webctrl aaname='Мадридский &quot;Реал&quot; стал самым дорогим футбольным клубом мира' tag='A' />

<html app='chrome.exe' title='Новости спорта, футбол, Лига чемпионов, УПЛ, бокс и другие виды спорта- 112.ua' />
<webctrl aaname='Мадридский &quot;Реал&quot; стал самым дорогим футбольным клубом мира' tag='A' />

```

Рисунок 3.6 - Приклад отриманої інформації із коду HTML – сторінки, який диференціює кожний об'єкт сторінки для програмного середовища Data Scraping

3.4 Стек використуваних ІТ, впровадження системи пошуку резюме за відкритими джерелами

Для розробки прикладної системи пошуку резюме за користувацькими критеріями, було використано програмне забезпечення UiPath, що є платформою для розробки прикладних рішень для взаємодії із користувацьким інтерфейсом, отримання та обробки інформації з нього. Також для отримання структурованих даних зі сторінки, було використано технологію Data Scraping.

UiPath — це глобальна компанія, що займається програмним забезпеченням, що виробляє програмне забезпечення для автоматизації роботизованих процесів (RPA). Він був заснований у Бухаресті, Румунія, Данієлом Дайнесом та Маріусом Тірке. Її штаб-квартира знаходиться в Нью-Йорку. Програмне забезпечення компанії відстежує діяльність користувачів для автоматизації повторюваних

завдань фронт- та бек-офіс, у тому числі виконуваних за допомогою іншого програмного забезпечення для бізнесу, такого як програмне забезпечення для управління відносинами з клієнтами або програмне забезпечення для планування ресурсів підприємства (ERP).

Написання часткового коду для аналізу відповідності критеріїв було зроблено на платформі .NET.

.NET Framework – програмна платформа, випущена компанією Microsoft у 2002 році. Основою платформи є загальномовне середовище виконання Common Language Runtime (CLR), яке підходить для різних мов програмування. Функціональні можливості CLR доступні у будь-яких мовах програмування, що використовують це середовище. В даний час .NET Framework розвивається як .NET.

Вважається, що платформа .NET Framework є відповіддю компанії Microsoft на велику популярність платформу Java компанії Sun Microsystems, яка набрала на той час (нині належить Oracle).

Хоча .NET Framework є патентованою технологією корпорації Microsoft і офіційно розрахована на роботу під операційними системами сімейства Windows, існують незалежні проекти (насамперед Mono і Portable.NET), що дозволяють запускати програми .NET Framework на деяких інших операційних системах.

Програма для .NET Framework, написана будь-якою мовою програмування, що підтримується, спочатку перекладається компілятором в єдиний для .NET проміжний байт-код Common Intermediate Language (CIL) (раніше називався Microsoft Intermediate Language, MSIL). У термінах .NET виходить збирання, англ. assembly. Потім код або виконується віртуальною машиною Common Language Runtime (CLR), або транслюється утилітою NGen.exe виконуваний код для конкретного цільового процесора. Використання віртуальної машини переважно, тому що позбавляє розробників необхідності дбати про особливості апаратної частини. У разі використання віртуальної машини CLR вбудований в неї JIT-компілятор "на льоту" (just in time) перетворює проміжний байт-код на

машинні коди потрібного процесора. Сучасна технологія динамічної компіляції дозволяє досягти високого рівня швидкодії. Віртуальна машина CLR також сама дбає про базову безпеку, управління пам'яттю і систему виключень, позбавляючи розробника частини роботи.



Рисунок 3.7 - Архітектура створення програми на платформі .NET

3.5 Інструкції користувача, адміністратора програми пошуку та структурування даних резюме із відкритих джерел

Структура прикладного застосунку для пошуку резюме за певними критеріями складається із двох структурованих частин:

Програмного коду та таблиці для вводу користувацьких критеріїв.

.entities	21.05.2022 11:27	Папка с файлами	
.local	08.06.2022 10:29	Папка с файлами	
.screenshots	22.05.2022 22:36	Папка с файлами	
.settings	21.05.2022 11:27	Папка с файлами	
.tmh	21.05.2022 11:27	Папка с файлами	
Job_Request.xlsx	23.05.2022 15:20	Лист Microsoft Ex...	11 КБ
Main.xaml	22.05.2022 22:37	UiPath Studio proj...	46 КБ
project.json	21.05.2022 11:27	JSON File	2 КБ

Рисунок 3.8 - Програмний склад розробленого застосунку

Файл для конфігурації користувацького запиту

	A	B	C	D	E	F	G	
1	#	Назва вакансії	Очікувана ЗП	Стать	Освіта	Зайнятість	Досвід	Результат
2	2	Дизайнер	20000	Чоловіча	Середня	Повна зайнятість	1	
3	3	.Net Developer	100000	Жіноча	Не закінчена вища	Критерій не важливий	3	
4	4	C++ Developer	45000	Не важливо	Спеціальна	Критерій не важливий	2	
5	5	Бухгалтер	25000	Жіноча	Вища	Повна зайнятість	2	
6								

Рисунок 3.9 - Структура таблиці користувачького запиту

Структура складається із 8 колонок, заповнюючи які користувач може конфігурувати майбутній запит.

Опис колонок таблиці представлено у таблиці 3.1

#	Порядковий номер елементу таблиці. Створюється автоматично при додаванні нового рядка, та є технічним інструментом реалізації застосунку
Назва вакансії	Назва пошукового запиту, який буде задано у архіви джерел з відкритих даних
Очікувана ЗП	Показник заробітної плати, вище якого кандидати будуть розглядатися із нижчим коефіцієнтом придатності до заданого запиту
Стать	Якщо важлива стать людини, яку шукає користувач, її можливо із викидного списку даної колонки. Якщо стать не важлива – стандартне значення колонки буде «Не важливо»
Освіта	Перелік можливих освітніх фахів, що було отримано із можливих критеріїв на ресурсах архівах відкритих даних
Зайнятість	Перелік можливих видів зайнятості, що було отримано із можливих критеріїв на ресурсах архівах

	відкритих даних
Досвід	Досвід кандидата, за яким прикладна програма має проводити пошук за джерелами відкритих даних
Результат	Поле у якому програма виведе перелік найпридатніших кандидатур за вибраними користувачем критеріями

Для початку роботи програми користувачу потрібно виконати наступний перелік дій:

1. Заповнити таблицю запитів інформацією, за якою потрібно знайти кандидата на вакантне місце роботи
2. Натиснути на виконавчий файл(Start_Search.bat) у теперішній деці для того щоб програма почала пошук.
3. Після закінчення роботи програми – користувач може отримати результат пошуку відкривши Excel таблицю у якій він вводив дані для пошуку та подивившись у колонку «Результат»

Програмний код застосунку, адаптований для підтримки платформою UiPath

Розділ проектів на менші робочі процеси є першорядним для коректного технічного дизайну проекту та є стандартом якості при розробці прикладних програмних програм. Файли робочого процесу розміром понад 10 МБ не мають бути створені оскільки мають занадто багато технічної інформації, що не є стандартом проектування прикладного застосунку.

Функціональний програмний код застосунку зберігається у файлах із розширенням .xaml та інтерпретуються у подальшому влаштованим у систему інтерпретатором.

Project.json — це автоматично згенерований файл, який створюється для кожного файлу .xaml, позначеного як Основний у папці проекту.

Файл містить різноманітну інформацію, включаючи залежності проекту або веб-служби, завантажені в бібліотеки.

Розглянемо найголовніші елементи даного файлу:

Параметр	Пояснення
Name	Назва проекту автоматизації. Він надається в UiPath Studio під час створення нового процесу або бібліотеки.
Description	Опис проекту. Він надається в полі Опис у UiPath Studio, коли створюється новий проект.
Main	Технічна точка старту проекту автоматизації. Він складається з файлу .xaml. Ім'я за замовчуванням – «Main.xaml». Він відображається як заголовок панелі конструктора, так і на панелі «Властивості» в UiPath Studio. Якщо користувач хоче спочатку виконати інший проект, його значення потрібно змінити на ім'я файлу .xaml, який потрібно обробити.

Файл Main автоматично створюється, та повинен містити основний робочий процес програмного продукту.

ВИСНОВКИ

За час виконання бакалаврської кваліфікаційної роботи було виконано усі поставлені перед роботою цілі, а саме:

- Отримані теоретичні засади щодо розроблення програмного забезпечення для отримання даних резюме з відкритих джерел
- Розроблений програмний продукт пошуку резюме за критеріями користувача

Для реалізації поставлених цілей проєктованій системі, було розглянуто та проаналізовано якісні існуючі програмні рішення та виділено функціонал, який найкраще підходить для створюваного застосунку та має мінімальний необхідний набір функцій для зручного використання користувачем, що не зменшує продуктивність роботи програми.

Проведений аналіз існуючих програмних рішень дозволив створити перелік найкращих практик для створення програмного забезпечення пошуку резюме кандидатів ґрунтуючись на архівах відкритих даних.

Під час виконання курсової роботи було проведено дослідження, якими було визначено найбільш влучні алгоритмічні рішення для забезпечення даної системи всім необхідним функціоналом, простотою розробки, тестуванням та зручною експлуатацією користувачем.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Електронний ресурс: .Net 4.7.2 2021 – Режим доступу: <https://docs.microsoft.com/en-us/dotnet/csharp> .
2. Електронний ресурс: Інформація щодо фреймворку UiPath REFramework - <https://www.uipath.com/community/rpa-community-blog/implement-uipath-projects-using-enhanced-reframework>.
3. Електронний ресурс: Офіційний сайт UiPath - <https://www.uipath.com/>.
4. Книга: Д.-С. Джеон та Н.Н. Есфакані, “Агрегатори новин та конкуренція серед газет в Інтернеті (попередні та неповні)”, 2012.
5. Книга: К. Лерман, “Обробка соціальної інформації при агрегуванні новин”, Інтернет-обчислення ІЕЕЕ, вип. 11, № 6, с. 16–28, 2007.
6. Книга: Н. Діакопулос, М. Де Чоудхурі та М. Неаман, “Пошук та оцінка джерел інформації в соціальних мережах у контексті журналістики”, у Матеріали конференції SIGCHI про людський фактор в обчислювальних системах, 2012, с. 2451–2460.
7. Книга: Р. Барзілай та М. Елхадад, “Використання лексичних ланцюжків для узагальнення тексту”, Досягнення автоматичного узагальнення тексту, С. 111–121, 1999.
8. Книга: С. Чоудхурі та М. Ландоні, “Послуги з агрегування новин: очікування та досвід користувачів”, Інтернет-огляд інформації, вип. 30, ні. 2, с. 100– 115, 2006.
9. Книга: Ф. Гамборг, Н. Меушке та Б. Гіпп, “Матрична агрегація новин: вивчення різних перспектив новин”, у Праці 17-го Спільна конференція АСМ / ІЕЕЕ з цифрових бібліотек. ІЕЕЕ Press, 2017, С. 69–78.
10. Стаття: Alok MT Learning Robotic Process Automation. Packt Publishing, 2018.
11. Стаття: С. Grozea, D.-С. Cercel, С. Onose і S. Trausan-Matu, “Atlas: Service aggregation news”, in2017 16-а конференція RoEduNet: Мережа в галузі освіти та досліджень (RoEduNet). ІЕЕЕ, 2017, с. 1–6